ISSN 1471-0498

DEPARTMENT OF ECONOMICS

DISCUSSION PAPER SERIES

TOWNS (AND VILLAGES); DEFINITIONS AND IMPLICATIONS IN A HISTORICAL SETTING

Florian Ploeckl

Number 536
February 2011

Manor Road Building, Oxford OX1 3UQ

Towns (and Villages); Definitions and Implications in a Historical Setting

Florian Ploeckl*

University of Oxford

Nuffield College

February 28, 2011

Abstract

Urbanization has been extensively used as a proxy for economic activity. The urban status of settlements is usually determined by an ad hoc population size threshold. This paper proposes a new threshold, taking into account the effect of local agricultural endowments. The new population threshold is a population size, such that for smaller settlements these endowments influence their size, while for larger they do not. This results in an endogeneous, data based threshold. The idea is practically shown for Saxony in the 19th century. The relevance of a different classification is demonstrated in four particular examples, the development of urbanization over time, Gibrat's law, the impact of geography on town locations and the spatial relationship between towns and villages. The results demonstrate that the underlying classification scheme matters for the conclusions drawn from urban data.

Keywords: Towns, Villages, Geography, Definition, Classification, Town Size

JEL Codes: N93, B49, O13, R12

---

*E-mail: florian.ploeckl@economics.ox.ac.uk

Nuffield College, 1 New Road, Oxford, OX11NF, United Kingdom

I want to thank Bob Allen, Rui Esteves and James Fenske for helpful discussions as well as seminar audiences at Oxford and conference audiences at the Sound Economic History and APHES meetings. All errors are of course my own.

1 Introduction

What is a town[1]? Although they are often used as the quintessential example for agglomeration economies, it is rarely addressed what exactly constitutes a town. The frequent use of such a concept obviously implies applied classification schemes, but the underlying definitions on which these are based are rarely made explicit. Pretty much nobody disputes that there is a distinction between rural and urban settlements, though the exact nature of this split is very much up in the air.

A formal distinction matters for a number of contexts. One of them is the use of towns and town based indicators, for example the urbanization rate of a region or state, by the economics and economic history literature. There they are utilized extensively as observations and proxies for other economic variables of interest, most notably general economic growth. Urban population growth has been used to determine the growth effects (or non-effect) of a whole range of different factors, examples range from governance structures (Long and Shleifer, 1993), Atlantic trade (Acemoglu, Johnson and Robinson, 2005), the invention of the printing press (Dittmar, 2010), the protestant reformation (Cantoni, 2009) or the introduction of the potato in Europe (Nunn and Qian, 2009). Many studies, including those mentioned above, use a simple definition of towns as settlements with a population larger than a particular size threshold. The choice of this approach is mostly justified by the availability of particular town size data sets. The actual choice of the size threshold is rarely addressed, usually it is simply taken over from the size threshold used in the data sources. Studies focusing on European history are normally based on the data sets by Bairoch, Batou and Chèvre (1988) and DeVries (1984), who use thresholds of 5000, respective 10000.

Growth is however just one example where the literature utilizes towns. In other cases the town size itself is the focus of attention. One example is the literature on the town size distribution looking at Zipf's law, which describes the seeming regularity of the distribution (Zipf, 1949; Gabaix, 1999). Explanatory approaches for this size rank relationship also include Gibrat's law, which postulates the independence between size and growth of a town (Gibrat, 1931; Eeckhout, 2004). Next to this statistical approach endowments and agglomeration effects are discussed as potential determinants for town size (Davis and Weinstein, 2002). Geographic endowments can influence the

[1] Although there are some regions where the terms town and city have slightly different connotations, I will use the terms interchangeably.

size of towns, but they can also influence where towns are located and thereby the spatial structure of the urban system (Weber, 1920; Christaller, 1933). The spatial distribution of towns is also at the center of the literature on agglomeration effects. Models of the New Economic Geographic stress the importance of market access for the spatial distribution of economic activity (Krugman, 1991; Fujita, Krugman and Venables, 1999). Through a focus on labor as input factor that also explains the spatial distribution of the population (Redding and Sturm, 2008). These relationships between locations however not only exist for towns, they also include the urban-rural connection between towns and villages (Duranton, 1998, 1999). All of these areas depend on a definition and connected practical classification of township for the respective conclusions.

There are a number of approaches to define towns, the most relevant in this context are economic ones, focusing on production, and institutional ones, for example based on legal rights. A second step is to turn the definition into a practical classification scheme. The use of an ad-hoc population size threshold is an example of how an economic definition is turned into a practical one (DeVries, 1984). Instead of developing a new conceptual approach I utilize the existing definition underlying the usual size threshold approach. I create a new classification scheme that turns this definition into a practical classification. I even retain the use of a population size threshold, though instead of using an ad hoc determination I use the importance of (non-)agricultural production, one of the criteria underlying the definition, to select a particular value. This importance is determined through the relevance of local agricultural endowments for settlement size, leading to a completely data based threshold.

After a more extended discussion of this idea I demonstrate its use in an actual example. The historical setting for my case study is Saxony between 1834 and 1871. This independent German state organized detailed population counts during this time. The published results are the basis of the dataset that contains all settlements, including small villages, within the state for the time period. Additionally detailed information about agricultural endowments are available for each location. The empirical tests result in a town threshold of 1650 inhabitants for 1834. This indicates that the usual size thresholds are neglecting to categorize a considerable number of towns and people as urban.

The second part of the paper details the implications of this approach by investigating the impact of different classifications on a number of areas. This is done using the three sets of Saxons towns defined by my new classification, a legal approach as well as the size threshold of 5000 inhabitants. The impact of using different township definitions is demonstrated in four particular examples, urbanization, Gibrat's law, the impact of geography on town locations and the spatial relationship between towns and villages. The results demonstrate that the underlying classification scheme matters for the conclusions drawn from town data.

The development of urbanization rates over time implies that the lessons about the relative economic growth of regions drawn from these rates should take into account whether urbanization can be derived identically in each region. Similarly, the validity of Gibrat's law for towns, the independence of a town's growth rate and size, is contingent on the underlying set of towns. The same holds for the influence of geographical factors on the location of towns, changing the classification scheme for towns does indeed change which endowments have a statistically significant influence on the emergence of towns at specific locations. The implications for the locations of towns also affect the relationship between villages and towns. The characteristics of the spatial patterns of villages and town locations differ depending on what settlements are seen as towns.

2 Urban definition

The existence and use of practical definitions of towns, mostly based on size cutoffs, does not imply that much attention is spent on an underlying conceptual definition. For example there is only a limited discussion about such a definition in urban economics textbooks or the Handbook of Urban economics (Henderson, Nijkamp and Thisse, 2004).[2] Economists focus much more on the size distribution, in particular on questions relating to Zipf's law, the noted empirical regularity about size and rank. Some of these papers look at the effect of threshold (Malecki, 1980; Guerin-Pace, 1995), however they do not link it to urban definitions, but focus on the implications for size-rank rules. There are a number of classifications used for example in the literature on the systems and hierarchy of locations (Christaller, 1933). Also regional science and the new economic geography deal with the different categories of locations, though the different categories are usually conceptual, assuming rural

[2]Historically this seems to be similar in other fields, for example Sociology (Martindale, 1958)

settlement to be dispersed and fully agricultural, while towns are concentrated and usually only produce manufacturing goods (Fujita, Krugman and Venables, 1999).

A separation into towns and villages, with the implicit definition of villages as those locations which are not urban, is also a strong element in the economic history literature on urban history. For example Bairoch in his work on the global history of cities and Devries in his work on European urbanization during the early modern period discuss characteristics of cities, which leaves all other settlements to be villages (Bairoch, 1988; DeVries, 1984). Their definitions follow in large parts earlier work on the definition of towns, in particular also the characterization laid out by Weber in his work on the concept of the city (Weber, 1920).

Broadly there are two relevant approaches to define towns and consequently villages.[3]These are economic ones, focusing predominantly on population and production structures of settlements, and institutional ones, focusing on legal and other formal institutions located within settlements.

One of the influential economic definitions is based on DeVries (1984). As mentioned, it takes up ideas developed by Weber (1920) and distills them into four practical, quantifiable criteria:

- Population size
- Share of non-agricultural population
- Diversity of non-agricultural occupational structure
- Density of settlement

All four of these are continuous indicators, so an empirical application needs to draw a line at some point for each of these to use this definition in a practical situation. There are no hard rules about what values are sufficient such that a location is called a town (DeVries, 1984). All of these are concerned with the nature of the production process at this location, focusing on population to measure its extent and spatial density. Two of them are more explicitly in this through their focus on the occupational structure, which is supposed to be neither dominated by agriculture nor any other industry.

The second approach utilizes institutional characteristics of settlements, most prominently a legal right of township. Such a status can be a purely formal characteristic without any particular real-world consequences or it can actually cause an

[3]There exist more than that, for example Tilly (1976) develops another one. It focuses on concepts of market structure and relationships.

institutional difference between towns and villages. Possible differences are in areas like governance structures, tax and fiscal issues, rights to trade fairs, or also military and security related. The existence of such a legal characteristic was especially pronounced in Europe's history. Among others, one of the major problems to utilize such a definition are structural differences between how different authorities applied the concept, so what passed for a town in England did not necessarily do so in German states. Unless there was a centrally applied concept, which usually implied a central government authority, a legal approach to define township faces severe issues to be applied in a consistent and useful way.

This discussion focused until now on a conceptual definition of township. The second component is to turn them into a formal classification scheme, in particular which settlement characteristics are used to identify locations as towns or villages. The economic definition by DeVries is usually reduced to the first criterium, population size. DeVries uses 10000, Bairoch uses 5000 as threshold for inclusion in the database of historical town sizes in Europe.[4] Both are rather ad hoc thresholds without a definite justification for the value. The main argument for the thresholds are that these values are high enough such that sufficient historical data can be found, but small enough that the error of misidentifying villages as towns is minimized. The legal definition appears to be more straightforward, settlements either have town rights or they don't. But this presumes that town rights are a distinct, unified set of rights. This was not always the case, locations with town rights might not have had all of them, while some villages had certain town rights, for example the right to hold trade fairs. A practical implementation therefore needs to select one particular legal or institutional characteristic. In the case of Saxony, the government reformed municipal governance in 1832 and classified locations directly into towns and villages. This codification is based primarily on historical settlement rights regarding governance and taxation, but these rights were not as clear as the distinction made by the government.

The definition underlying my classification is based on the criteria described by DeVries (1984). The four central ones are population size, share of non-agricultural population, diversity of non-agricultural occupational structure and density of settlement. The usual ad hoc size based classifications imply that the latter three criteria are not taken into account. My new classification will improve on this by incorporating information on the importance of agriculture, therefore explicitly covering the

[4] They do attempt to provide the size of towns for the time before they reach the thresholds.

second criteria as well. Both of these measures are continuous, so they suffer from the same threshold problem. There is no clear exogenously given value which separates towns from villages. Furthermore the share of non-agricultural population is often not as readily available as other data, in particular population size.

These issues are overcome in two steps. First, instead of using a sectoral population share, which requires essentially an occupational census, I utilize local agricultural endowments. These are quite consistent over time and usually cover the complete area in question, which implies better data availability. Second, instead of using two separate criteria, I drop the sectoral share of agriculture as an independent characteristic and use the underlying idea about the relevance of agriculture for town income to derive a data-based classification threshold for population size.

Towns are assumed to have a high population and non-agricultural occupations while villages are small and are more strongly agricultural. Instead of drawing two arbitrary lines I draw a population threshold based on the non-agricultural production criterium. To turn the importance of this factor into a selection procedure for a population size threshold I utilize the relevance of agricultural endowments. These endowments influence extent, structure and characteristics of local agriculture. Consequently, if a settlement is a village, therefore primarily agricultural, then the local endowments can be expected to have an impact on the settlement. Towns in contrast are predominantly non-agricultural, so they should be unaffected by and independent of the local conditions for agriculture. This distinction between urban and rural areas is similar to that used by the new economic geography (Fujita, Krugman and Venables, 1999). There it is the result of modelling choice, that defines villages and rural areas as producers of agricultural goods, while towns are assumed to produce non-agricultural goods. Here I am focusing on the implications for income and consequently sustenance for population.

The main assumption is the importance of local endowments. Local here implies the direct vicinity rather than the larger hinterland region. The local endowments should characterize the land which is used by the village or town farmers for their production. This is necessary to ensure a link between the characteristics of these local endowments with the relevance of agricultural production for the income of the settlement population. The endowments essentially determine a constraint for the amount of population a village can support. Technological change might change that over time, but as long as the constraint relationship is fairly equal across the state the approach underlying the definition holds.

If towns specialize in the production of non-agricultural goods, it obviously implies that they rely on trade with the countryside to provide the required food supply.[5] If this relationship between a town and its hinterland is assumed to be closed, like an island, than the regional agricultural conditions influence town population through the production constraint. The regional conditions are not necessary correlated with the local conditions of the town. The number of urban residents is restricted by the amount of food the hinterland can produce. This idea can be scaled upwards to apply to whole urban systems and is used to measure agricultural productivity of countries by the number of persons one agricultural worker supplies food for (Wrigley, 1985; Allen, 2000). While this is a worthwhile approach for whole states, for individual towns even this regional constraint is not really binding. The reason for this is trade between towns, which is relatively seen far more extensive than food exports and imports between states. Although agriculture might provide a constraint for the total urban population in a state, internal trade between settlements can lead a distribution of the urban population that does not correspond to local agricultural conditions. Furthermore other constraints on a town's population, for example the impact of density on health and therefore mortality, might bind, which implies that the agricultural structure is not the relevant factor for determining a town's size. A further constraint was the ability of towns to restrict entry and therefore size through institutional and economic restrictions. One influential historical example for such restrictions was the influence of guilds on the economic life of towns and thereby their size (Duranton, 1999).

2.1 Practical Implementation

The above discussion details the conceptual idea of the new definition and the process to turn it into a practical method. The following section applies this method to a historical case. The chosen setting is Saxony in the early 19th century. This example will practically demonstrate how this method leads to a new size threshold.[6]

[5]The production of non-agricultural goods does include food processing just not the initial production of food inputs.

[6]The specific relevance of the two criteria that are not used in the classification process is discussed in the appendix

2.2 Saxony 1834

The main data required to determine the new threshold is the full size distribution, covering all locations within the state. I utilize a recently collected data set[7], which covers the full set for Saxony in the middle of the 19th century. In particular the set contains the years 1834, 1843, 1852, 1861 and 1871, covering the time period between Saxony's entry into the Zollverein, the 1834 customs union between German states, and the loss of its independence through becoming part of the new German empire founded in 1871. Saxony was a mid-sized German state with a long history. The central territories, which are part of this dataset, were under its control for about four centuries, in parts even longer. It was a center of international trade, controlled a number of international trade routes and had a major international trade fair at the city of Leipzig. Saxony was one of the early industrializing German states, entering the industrial revolution by the time in question (Kiesewetter, 2007; Forberger, 1982). Its economic structure was fairly diverse, in 1849 it had the following distribution. Agriculture employed 32% of the population, crafts and manufacturing 51%, trade 3%, transport 2%, art and science related occupations 4% and household servants 3% (Bureau, 1854). Saxony was relatively densely settled, the population density was one of the highest among German as well as European states (Kiesewetter, 2007). The main settlement process had started by the 14th century, by the 16th the set of settlement locations was quite stable (Blaschke, 1967). Furthermore local governance structures evolved in such way that the legal institution of township had emerged, which gave a number of settlements specific institutional structures and rights. Those structures affected town governance and economic regulations, for example taxation, though their importance faded during the 19th century and was essentially meaningless after the state was absorbed into the empire in 1871. The existence of these legal structures allows the comparison of size based approaches with an institutional one. Furthermore the nature of its settlement process was such that individual homesteads clustered strongly into villages and towns; settlements had therefore clear boundaries. There were a number of different village forms, but all of them cluster people's houses rather than locating them on individual farming land. The conclusion of the process some two centuries before the time in question implies that there were no significant parts of the state which were not settled or under some form of cultivation (Blaschke, 1967).

[7] The set is based on data used in Ploeckl (2010*b*)

Although elements of the Saxon tax system allow for the calculation of total population as well as for some of the settlements, the first modern population count was conducted in 1834 as a consequence of Saxony's entry into the Zollverein, the customs union with other German states.[8] These counts were conducted in a three year rhythm until the creation of the empire in 1871.[9] At the turn of the 20th century officials of the Saxon statistical office published short histories of Saxon towns and villages in the 19th century including all available population data (Waechter, 1901; Lommatzsch, 1905). This is the basis for the population part of the dataset. In total the set contains 3579 locations with a total population of 1.60 million in 1834, which increased to 2.56 million in 1871. The location population ranged in 1834 from six to 73610 with an average of 447 and a median of 201. By 1871 this had changed to a range from seven to 177100 with an average of 716 and a median of 261.

Each location is referenced with geographic coordinates[10], usually of a central spot within the settlement., which allows to link it with a number of geographic data, agricultural and otherwise. The main component is information about the aptitude of local land for agricultural and pastoral purposes. These are based on extensive field surveys undertaken by the Saxon government (Blaschke and Klasse, 1998). The survey combined various soil, water and climate indicators to derive a map of land quality with regard to farming as well as pasture purposes. This original information was used by the Saxon authorities to create an average value for each current municipality. There are about 1600 of those, which implies that each observation combines the average for just over two villages. The survey classifies all locations on a scale from zero to 100, reporting separate numbers for farming and pasture purposes.

2.3 Distributional characteristics

The classification of towns and villages implicitly assumes that the complete distribution of settlements can be separated into at least two sections. A pure size-based definition obviously simply cuts the size distribution at a particular size value, while

[8]The revenue distribution scheme of the Zollverein was based on a state's population, the states therefore agreed to consistent methods for population counts (Hahn, 1984; Henderson, 1984; Ploeckl, 2010*a*).

[9]The German Empire continued to conduct counts in a five year rhythm.

[10]These are either official coordinates from the Saxon *Landesvermessungsamt*, from a historical place register or selected by the author by inspecting various maps (Blaschke and Baudisch, 2006).

one based on certain town characteristics, for example legal status, does not exclude the possibility that there are towns which are smaller than other settlements which do not have corresponding rights. Nevertheless in both cases the full set of settlements is separated into two distinct ones.[11] Such a break does not necessarily go against approaches which look at the distribution as a whole (Eeckhout, 2004), though it fits better with approaches that focus on a subset of the locations (Levy, 2009). Two of the main arguments Eeckhout (2009) presents in a reply are the statistical conformity of the empirical distribution and the indeterminacy of the breakpoint. The selection of the size threshold based on agricultural endowments creates a data-based breakpoint, thereby addressing the second point. Before doing so I shortly investigate the first issue. Eeckhout uses modern US data when he demonstrates that the size distribution of all settlements fits a log-normal distribution. I test the same for the full set of Saxon settlements for a number of years between 1834 and 1871. In all cases the hypothesis can be rejected with a 99% confidence level. This holds for a number of different test approaches.[12] This rejection implies that the size distribution of Saxon settlements did not follow a log-normal distribution, which negates the applicability of the underlying theoretical approach of Eeckhout (2004). This points towards the suitability of an approach that segments the total distribution into at least two classes.

2.4 Determining the threshold

Taking up the idea developed above I apply this classification approach to the practical example of Saxony. The basis of this test is the following empirical specification

$$\begin{aligned}&\ln(Pop) = \\ &\alpha + \beta_f Farm + \beta_p Pasture + \beta_{fp} Farm * Pasture + \delta * Large_t + \\ &+ \gamma_f Farm_t * Large_t + \gamma_p Pasture * Large_t + \gamma_{fp} Farm * Pasture * Large_t + \epsilon\end{aligned}$$

where Pop is the size of each settlement in the particular year in question and Farm and Pasture represent the location's soil quality with regard to farming and pasture purposes. $Large_t$ is a dummy variable indicating whether the location in question

[11] Studies have highlighted the special role of the largest town, often the capital (Ades and Glaeser, 1995).Since this concerns only one specific town, the impact is limited and therefore negligible in this context.

[12] The applied tests were Lillie, Shapiro and K-S.

has a population size larger than a certain size threshold t. The estimation is done separately for a range of possible size threshold values.

This specification allows me to test for which size threshold t the influence of local agricultural endowments on the size of the settlement changes.[13] This is formally done by using an F-test for the joint hypothesis that $\beta_i + \gamma_i = 0$ for $i = f, p, fp$. If the hypothesis is not rejected the test indicates that the combined effect of the agricultural variables on the size of settlements below the threshold is offset for those settlements above the line. The test in itself only indicates whether the combined effects offset each other, to distinguish whether there is an effect below the threshold, above the threshold or in neither of the two an additional test is necessary. If the hypothesis of the offset of the joint effect cannot be rejected, a F-test rejecting the joint hypothesis that $\beta_i = 0$, for $i = f, p, fp$ is sufficient to show that there was an effect of agricultural variables below the threshold but not above.

The estimation of the specification and the hypothesis tests are repeated for a range of threshold size values. I test for the full range for values from 1000 to 5000 in steps of 10.[14] Using this rather fine incremental change in the threshold allows me to determine the largest settlement size for which the two mentioned tests, the existence of an effect below the threshold and the offset above, quite precise.

Figure 1 shows the results graphically using Saxony's population data for 1834. It plots the p-value of the offset test against the respective threshold value. As marked in the plot this leads to the determination of a threshold of 1650 inhabitants.[15] The test about the effect of agricultural variables on the size of settlements below the threshold rejects the hypothesis of no effect with a 99% significance level. These results imply that in 1834 locations below 1650 inhabitants show a statistically significant influence of agricultural endowment variables, while those above show no such influence.

The results show that an agricultural endowment based definition of towns and villages implies a comparatively low size threshold. In the case of Saxony in the early 19th century this classification scheme has as a result that 154 settlements, which represent 4.4% of all locations, can be considered to be towns. The average size of these towns is 3933, with a median of 2378. With a combined population of 605618 this implies a degree of urbanization of 38.7%.

The resulting threshold is a credible number, though surely pushing towards the

[13]Table 1 establishes that endowments have an effect on the size of all villages.

[14]As a robustness test I also tested for values up to 10000, the results did not change.

[15]The selection was based on a p-value of 0.10.

lower limit of what conceivably can be a reasonable value. This also implies that the relative size of the type of error committed changes. One of the reasons behind the selection of the ad-hoc threshold is to avoid type II errors, namely falsely identifying a settlement as a town when it fact it is a village (DeVries, 1984). This obviously neglects type I errors, namely failing to identify settlements as towns when they actually are. The resulting new threshold clear shifts the focus towards a more balanced view on the two errors

3 Implications

The new definition results in a considerable different population threshold than commonly used to classify towns. But does it actually matter? The following section explores the implications of different definitions by investigating a number of economic relationships commonly deduced with the help of urban data. The first is the development of urbanization over time. The second looks at the growth of towns. The literature on the town size distribution shows that Gibrat's law, the independence of relative growth and size of a town, can explain the existence of Zipf's law. I test whether Gibrat's law holds for the different sets of towns. The third concerns the geographic determinants of the spatial distribution of towns. Location characteristics might influence the likelihood that a town emerges on a specific location within the state. I test whether the same geographic factors influence this likelihood under the different definitions. The last test continues the issue of the spatial distribution but focuses on the interdependence of towns and villages. These tests are not designed to prove that one way to define towns is necessarily better than another. They are simply intended to demonstrate whether changing the underlying classification scheme really has an impact.

3.1 Urbanization

Saxony was one of the first regions in central Europe to begin the industrialization process. The main time frame for this process is the early 19th century (Kiesewetter, 2007; Forberger, 1982). The onset of the industrial revolution in a region had usually strong effects on the development of its urban centers, a prime example is Lancashire in England. Although there was no Saxon town which reached the size of these large urban agglomerations, they were recognized as major centers of industrialization within German territories. Chemnitz's nickname as the "Saxon Manchester" is a prime example for this.

The measured degree of urbanization is obviously impacted by the definition of township. I compare the rates based on different classifications to illustrate the impact of the underlying definition. This comparison looks not only at the 1834 levels but also on the development in the following decades. The basis are the above described data covering a number of years between 1834 and 1871. The same empirical approach is used to determine the set of towns implied by the population distribution and agricultural endowments for each of these years, the implied threshold, derived from the results shown in Figure 2, moves from 1650 in 1834 to 1750 in 1843, 2040 in 1852, 1880 in 1861 and 2180 in 1871. The increase in the size threshold over the course of the industrial revolution might be the result of increasing specialization between the agricultural dominated countryside and industrializing urban areas.

Furthermore I compare the resulting degree of urbanization to that derived from the size threshold of 5000 inhabitants and the legal definition. Figure 3 plots the calculated development of urbanization in Saxony for the industrial revolution using the different concepts.

Given the size threshold of 1650, the new measure obviously results in the highest degree of urbanization in 1834. The level of difference to the legal definition is small but noticeable. The difference to the 5000 threshold is however quite sizeable, which implies that the share of the population in locations between the new and the 5000 threshold is relatively large, more than double the share of the population residing in towns above 5000 people. The graph also shows that the share of the population in large, i.e. above 5000 inhabitants, towns grew significantly faster than the share of smaller towns. This indicates that the industrialization process was more pronounced and stronger in larger towns. Furthermore the share of people living in legal towns is growing considerably slower than the shares calculated based on any of the size definitions. This shows that the selection based on a legal definition is missing out a number of locations which were among the fastest growing during the industrial revolution. The static nature of this definition, the set of towns is essentially the same over the time period, does not adjust to the structural economic changes very well. A final point is the shape of the graphs, only the new approach shows a slowdown of the increase in the time period at the end of the 1840's. A possible explanation is the ability of this definition to react to the effect of the crisis in the last years of this decade due to a number of bad harvests, which hampered further industrialization and urban growth.

3.2 Growth

The growth of the town population is the first illustration of the impact the different definitions make. A major strand of the literature on town growth focuses on the statistical properties of the size distribution, in particular Zipf's law. A central explanation for the emergence of this regularity is Gibrat's law, which postulates that the size and growth of a town are independent. My first test is to investigate whether this proportionate growth hypothesis holds for the different sets of settlements classified as towns. The time frame of this analysis is the period from 1834 to 1871. This strikes a balance between a sufficiently long period to smooth out the impact of short-term fluctuations and ends early enough to avoid the issue of incorporations of villages and small towns into larger cities. The following specification is estimated:

$$\frac{Pop_{1871}}{Pop_{1834}} = \alpha + Pop_{1834} + \epsilon$$

Table 2 shows the results for the three town definitions, illustrating the differences between the newly developed one with the legal as well as the 5000 inhabitants threshold one. The hypothesis of proportional growth can be rejected for the towns classified with the new definition or the legal one. It cannot be rejected for towns that have more than 5000 inhabitants in 1834. This difference illustrates that the acceptance or rejection of hypothesis is contingent on the data sample and principles informing its creation. It also indicates that a simple separation of the whole distribution into two categories could potentially be improved by the introduction of more categories.

3.3 Geography

Geographic endowments are one of the main explanatory factors for urban size and growth. These factors are however also seen as factors for the emergence, and therefore location, of towns. Here I will demonstrate whether changing the definition of township will make a difference for the relevance of geographic endowments for town locations. In particular I test how these factors influence the likelihood that a particular site in Saxony is settled with a town.

The investigation is carried out using a spatial point pattern approach, which starts out with a set of locations, irregularly distributed within a region, like towns within a country, and assumes it to be generated by some unknown random mechanism (Diggle, 2003). The observed pattern x will be treated as the realization of a random point process X, where the number of points as well as the point locations in the two-dimensional region W are random(Baddeley and Turner, 2006). The interest is then into the parameters of the process X including the effect of explanatory variables like geographic factors. Of particular interest is the *intensity* of the point process, which is the expected number of points per unit area.[16] $\lambda(u)$ is the *intensity function*, which satisfies $E[N(X \cap B)] = \int_B \lambda(u)du$ for all regions B, assuming that $\lambda(u)du$ is the equal to the expected number of points falling in a small region of area du around a location u (Baddeley and Turner, 2006).

The impact of local characteristics is modelled through the influence of covariates on the intensity function, These covariates are based on spatial functions $Z(u)$ that are potentially observable at every spatial location $u \in W$. The intensity function $\lambda_\theta(u)$ now depends on a parameter θ, which leads to the following log-likelihood for θ:

$$\log L(\theta, x) = \sum_{i=1}^{n} \log \lambda_\theta(x_i) - \int_W \lambda_\theta(u)du$$

which is a well-behaved likelihood, but the MLE $\hat{\theta}$ is not analytically tractable and requires a numerical solution (Baddeley and Turner, 2006). Berman and Turner (1992) develop an algorithm that uses a formal similarity between the Poisson log-likelihood and that of a loglinear Poisson regression. This requires that the intensity function $\lambda_\theta(u)$ is loglinear in the parameter θ, formally $log\lambda_\theta(u) = \theta * S(u)$, where $S(u)$ is a real-valued or vector-valued function of location u. In particular $S(u)$ can be a function of observed spatial covariates. This leads to the use of the following form of the intensity function:

$$\lambda(u) = \exp(\alpha + Z\beta)$$

where α is a constant and Z is a vector of spatial covariates. I test for the impact of geographic factors using the distribution of towns based on three definitions, my new one, the legal definition and a size threshold of 5000. The set of geographic factors includes the agricultural factors described above and as well as elevation, ruggedness, temperature, rain, proximity to rivers and the Elbe, the main navigable river, and the distance to major and minor roads.

[16]If the intensity is constant over all of W it is referred to as uniform or homogeneous, while it is labelled inhomogeneous if it varies from location to location (Baddeley and Turner, 2006).

Table 3 shows the different impact of geographic factors on the likelihood that a particular location in the state is the site of a town. The three columns indicate which location characteristics have a statistically significant effect for towns based on the three different definitions, my new one, legal towns and towns with more than 5000 inhabitants. Since there is a considerable overlap between the different sets of towns, it is not surprising that two explanatory factors are statistically significant for all three, namely elevation and rain fall. The coefficients are such that at mean elevation levels the likelihood for a legal town is twice that for a town under the new definition and about 25 times that of a town with more than 5000 inhabitants. This is reversed for rainfall, where at average rain levels legal towns are 1/800 times and new definition towns 1/40 as likely. The main distinction between the definitions however is which of of the explanatory factors are statistically significant. While towns over 5000 inhabitants only show a significant influence of elevation and rainfall, legal towns additionally show the influence of rivers as well as the Elbe, the only navigable river in the area. The likelihood for towns under my new definition shows statistically significance for all the same factors as legal towns but differs through the relevance of even more factors. The levels of temperature as well as the distance to major roads have a statistically significant influence on the likelihood that a particular location is settled with a town.

Geographic factors influence the likelihood that particular location becomes the site of a town. The results also clearly demonstrate that the different definitions of township have a strong impact on the identification of the relevant geographic factors and the magnitude of their impact.

3.4 Spatial Pattern

Agglomeration effects are the other major explanatory factor for urban size. Market access often includes not only the town itself but also settlements in the vicinity. This leads to the question about the importance of township definitions for the spatial patterns of towns and villages.

The set of all locations is again treated as a spatial point process as above, but each point is assigned an additional piece of information, usually referred to as a its mark. A marked point pattern is therefore an unordered set $y = (x_1, m_1), ..., (x_n, m_n), x_i \in W, m_i \in m$, Here the information contained in a point's mark m_i is its town status, a simple 0/1 indicator.

The first hypothesis tested here is the spatial relationship between towns and villages. Do villages cluster around towns or are they dispersed away? Testing this hypothesis is based on nearest neighbour functions, which analyze the distribution of pairwise distances between observed points. The approach was introduced by Ripley (1977) with the specification of his K-function. The basic idea is the $\lambda K(r)$ is the expected number of other points of the process within a distance r of a typical point of the process (Baddeley and Turner, 2006). This can be adjusted for marked patterns such that $K_{ij}(r)$ is $1/\lambda_j$ times the expected number of points of type j within a distance r of a typical point of type i. To ease interpretation the following transformation is used: $g_{ij}(r) = \frac{K'_{ij}(r)}{2\pi r}$ Under the null hypothesis that X_i and X_j, the separate point processes for towns and villages, are independent, then $g_{ij} = 1$. Values $g_{ij}(r) > 1$ suggest clustering or attraction at distance r, so villages cluster around towns, while values $g_{ij}(r) < 1$ suggest inhibition, villages are actually dispersed away from towns.

Figure 4 plots the functions for the three definitions. it also plots the results of a number of Monte Carlo simulations of the underlying null hypothesis (Ripley, 1981). The evidence to reject the hypothesis that towns and villages are independently distributed is rather weak in all cases. Noticeable differences between the three definitions are such that the new definition exhibits a more dispersed pattern for short distances, the legal towns and villages seem essentially uncorrelated and towns based on a larger threshold appear to have a slight clustering effect for closer distances.

The previous test looks at the relationship between towns and villages as if they are completely different entities. The second tested hypothesis, called *random labelling*, starts out by interpreting all the locations as given and then looks at the spatial distribution of towns within the total settlement pattern (Baddeley and Turner, 2006). It interprets the emergence of an urban system as a two stage process, first the sites of all settlements are selected and then a number of settlements emerge as towns. The null states that whether a location becomes a town is independent of the status of all other locations. The test is again based on Ripley's K-function. The utilized statistic shows the difference between the summary function $K_{i\bullet}$, focused around the towns in the marked process, and K, the statistic for the point process without marks. If the statistic is positive then locations are more clustered around towns then expected given the total observed location pattern.

Figure 5 plots the resulting function values, again combined with a band based on

Monte Carlo simulations. The results indicate that for the set of towns based on the new as well as the legal definition the distribution of towns within the set of locations is not random. The same cannot be said for the set of large towns. The deviations from randomness for the new/legal definitions point towards a more regularly spaced pattern of settlements around towns.

As in the previous sections, the results of the analysis demonstrate that classifications matter. The role of towns within the spatial pattern of locations, especially the relationship between towns and villages, differs depending on what settlements are identified as towns.

4 Conclusion

Towns, either as direct object of study or as indicator for other economic characteristics, have long been the subject of interest for economists and economic historians. Empirical analysis however requires a classification scheme that defines what exactly constitutes a town. Numerous studies, especially on European history use ad hoc size thresholds, especially the number of 5000 inhabitants. However authors who compiled historical population data for Europe indicated that this likely misses out on a considerable number of urban areas with population numbers below this size. This paper uses the case of one European state, Saxony, to empirically test for an appropriate size boundary utilizing the advantages of a historical setting. The underlying idea to define the threshold is the (missing) impact of agricultural endowments on settlement size. The results suggest that a lower boundary, in the vicinity of 2000 inhabitants, reflects the conceptual definition underlying the separation into towns and villages more precisely. The results furthermore suggest that the boundary is not static, but does vary over time. The new definition admittedly has higher data requirements, but it demonstrates that the use of a simple, high and common threshold for multiple areas neglects not only a considerable part of the total population, but also misses adjustments for different sizes at which urbanization starts in different regions.[17] Next to urbanization three more tests demonstrate that the choice of a different classification has an impact on empirical relationships based on urban data. The empirical validity of Gibrat's law, the relevant geographic characteristics for the

[17]The appendix demonstrates a slightly different formal specification of the test, which requires less data. The results differ only slightly from the ones of the main analysis.

emergence of urban locations and the spatial relationship between villages and towns demonstrate that studies using urban data need to address the ramifications of their choice of an urban definition. The empirical implications however also offer opportunities to develop urban definitions and classification schemes based on the ideas and concepts underlying the theories tested with urban data.

References

Acemoglu, D., S. Johnson, and J. Robinson. 2005. "The Rise of Europe: Atlantic Trade, Institutional Change and Growth." *American Economic Review*, 95(3): 546–579.

Ades, A. F., and E. L. Glaeser. 1995. "Trade and circuses: explaining urban giants." *The Quarterly Journal of Economics*, 110(1): 195–227.

Allen, R. C. 2000. "Economic structure and agricultural productivity in Europe, 1300–1800." *European Review of Economic History*, 4(01): 1–25.

Baddeley, A., and R. Turner. 2006. "Modelling spatial point patterns in R." *Case studies in spatial point process modeling*, 23–74.

Bairoch, Paul. 1988. *Cities and economic development : from the dawn of history to the present.* Chicago:University of Chicago Press.

Bairoch, P., J. Batou, and P. Chèvre. 1988. *Population des villes européennes de 800 à 1850.* Droz.

Berman, M., and T. R. Turner. 1992. "Approximating point process likelihoods with GLIM." *Applied Statistics*, 41(1): 31–38.

Blaschke, K. 1967. *Bevölkerungsgeschichte von Sachsen bis zur industriellen Revolution.* Weimar,:Böhlau.

Blaschke, K., and P. H. Klasse. 1998. *Atlas zur Geschichte und Landeskunde von Sachsen.* Verl. der Sächsischen Akademie der Wissenschaften.

Blaschke, K., and S. Baudisch. 2006. *Historisches Ortsverzeichnis von Sachsen.* Leipziger Universitätsverlag.

Bureau, Statistisches. 1854. "Die Bevoelkerung des Koenigreichs nach Berufs- und Erwerbsclassen und Resultate der Gewerbs-Geographie und Gewerbs-Statistik von Sachsen." *Statistische Mittheilungen aus dem Koenigreich Sachsen*, 3.

Cantoni, D. 2009. "The economic effects of the Protestant Reformation: Testing the Weber hypothesis in the German lands." *URL: www. people. fas. harvard. edu/˜ cantoni/cantoni_jmp_2_7_1. pdf [16.03. 2010].*

Christaller, W. 1933. *Die zentralen orte in Süddeutschland.* Jena.

Davis, D.R., and D.E. Weinstein. 2002. "Bones, Bombs, and Break Points: The Geography of Economic Activity." *The American Economic Review*, 92(5): 1269–1289.

DeVries, Jan. 1984. *European urbanization, 1500-1800.* Cambridge, Mass.:Harvard University Press.

Diggle, P. J. 2003. "Statistical analysis of spatial point patterns."

Dittmar, J. 2010. "Information Technology and Economic Change: The Impact of the Printing Press." Working Paper.

Duranton, G. 1998. "Labor specialization, transport costs, and city size." *Journal of Regional Science*, 38(4): 553–573.

Duranton, G. 1999. "Distance, land, and proximity: economic analysis and the evolution of cities." *Environment and Planning A*, 31(12): 2169–2188.

Eeckhout, J. 2004. "Gibrat's law for (All) cities." *American Economic Review*, 94(5): 1429–1451.

Eeckhout, J. 2009. "Gibrat's law for (all) cities: reply." *The American Economic Review*, 99(4): 1676–1683.

Forberger, R. 1982. *Die industrielle Revolution in Sachsen 1800-1861.* Stuttgart:Steiner Verlag.

Fujita, M., P.R. Krugman, and A. Venables. 1999. *The spatial economy : cities, regions and international trade.* Cambridge, Mass. ;aLondon:MIT Press.

Gabaix, X. 1999. "Zipf's Law and the Growth of Cities." *The American Economic Review*, 89(2): 129–132.

Gibrat, R. 1931. "Les inégalités économiques." *Librairie du Recueil Sirey, Paris.*

Guerin-Pace, F. 1995. "Rank-size distribution and the process of urban growth." *Urban Studies*, 32(3): 551–562.

Hahn, Hans-Werner. 1984. *Geschichte des Deutschen Zollvereins.* Goettingen:Vandenhoeck und Ruprecht.

Henderson, J. V., P. Nijkamp, and J. F. Thisse. 2004. *Handbook of Regional and Urban Economics: Cities and Geography.* Elsevier.

Henderson, W.O. 1984. *The Zollverein.* Vol. 3rd, London, Eng. ; Totowa, N.J.:F. Cass.

Kiesewetter, Hubert. 2007. *Die Industrialisierung Sachsens : ein regional-vergleichendes Erklaerungsmodell.* Stuttgart:Franz Steiner Verlag.

Krugman, P. 1991. "Increasing returns and economic geography." *Journal of political economy*, 99(3): 483–499.

Levy, M. 2009. "Gibrat's Law for (All) Cities: Comment." *The American Economic Review*, 99(4): 1672–1675.

Lommatzsch, Georg. 1905. "Die Einwohnerzahlen der Landgemeinden von 1834 bis 1900 und die Veraenderungen in der Verwaltungseinteilung des Koenigreiches seit 1815." *Zeitschrift des koeniglichen saechsischen statistischen Landesamtes*, 51(1): 12–91.

Long, J. B De, and A. Shleifer. 1993. "Princes and Merchants: City Growth Before the Industrial Revolution." *Journal of Law and Economics*, 36(2): 671–702.

Malecki, E. J. 1980. "Growth and change in the analysis of rank-size distributions: empirical findings." *Environment and Planning A*, 12(1): 41–52.

Martindale, D. 1958. "Prefatory remarks: The theory of the city." *Max Weber, The City*, 9–67.

Nunn, N., and N. Qian. 2009. "The potato's contribution to population and urbanization: Evidence from an historical experiment." *NBER Working Paper.*

Ploeckl, F. 2010*a*. "The Zollverein and the formation of a customs union." *Oxford University Economic and Social History Series.*

Ploeckl, Florian. 2010*b*. "Borders, Market Size and Urban Growth, the Case of Saxon Towns and the Zollverein in the 19th Century." *Institut d'Economia de Barcelona Working Paper*, 966.

Redding, S.J., and D.M. Sturm. 2008. "The Costs of Remoteness: Evidence from German Division and Reunification." *American Economic Review.*

Ripley, B. D. 1977. "Modelling spatial patterns." *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(2): 172–212.

Ripley, B. D. 1981. *Spatial Statistics. 1981.* Wiley, New York.

Tilly, C. 1976. *Vendee: A Sociological Analysis of the Counter-Revolution of 1793.* Harvard University Press.

Waechter, Georg. 1901. "Die saechsischen Staedte im 19.Jahrhundert." *Zeitschrift des koeniglichen saechsischen statistischen Landesamtes*, 47(1): 179–232.

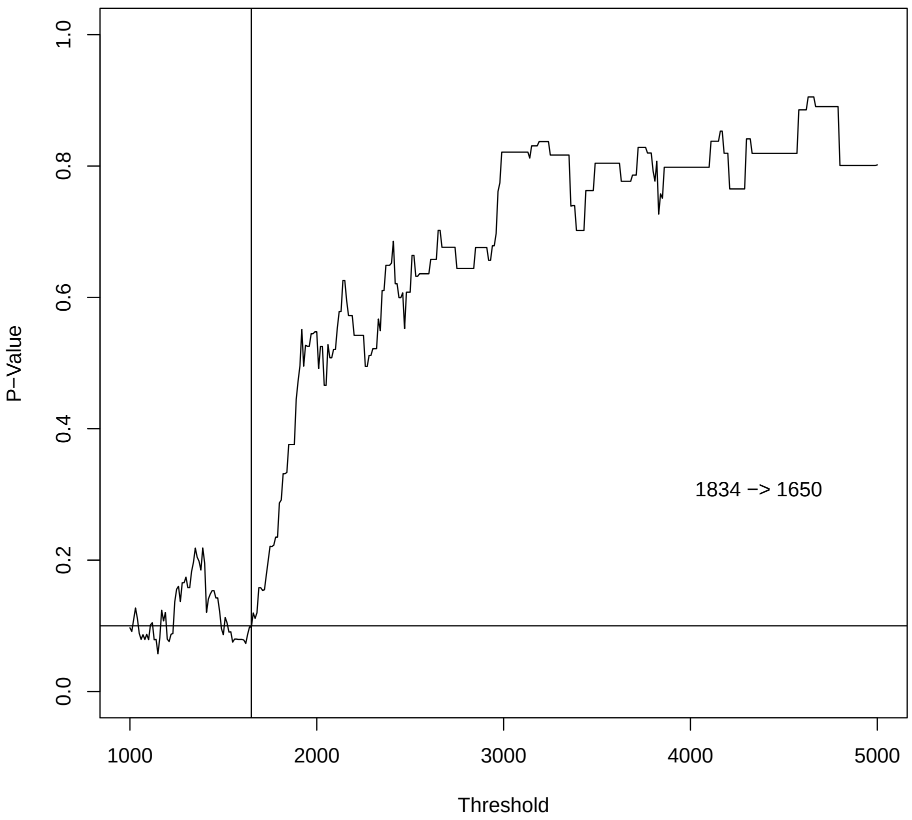
Weber, M. 1920. *Die Stadt.* Mohr.

Wrigley, E. A. 1985. "Urban growth and agricultural change: England and the continent in the early modern period." *Journal of Interdisciplinary History*, 683–728.

Zipf, G. K. 1949. *Human Behavior and the Principle of Least E ort.* Addison-Wesley Press, Cambridge, Massachusetts.

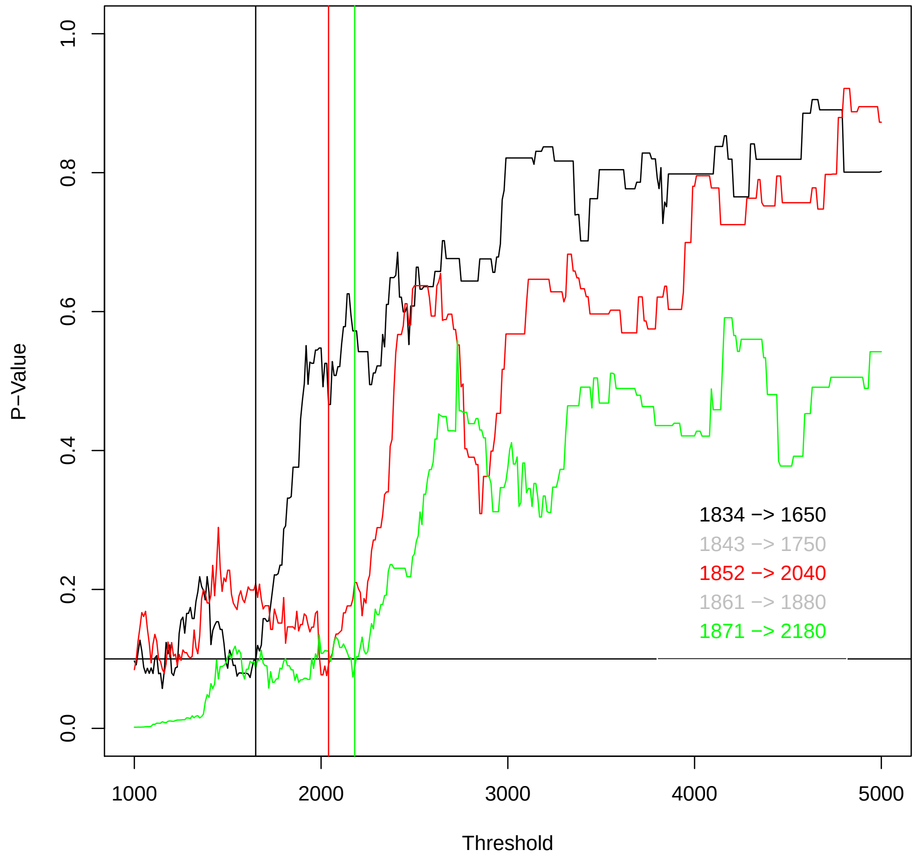
Figures

Figure 1: Threshold determination Saxony 1834



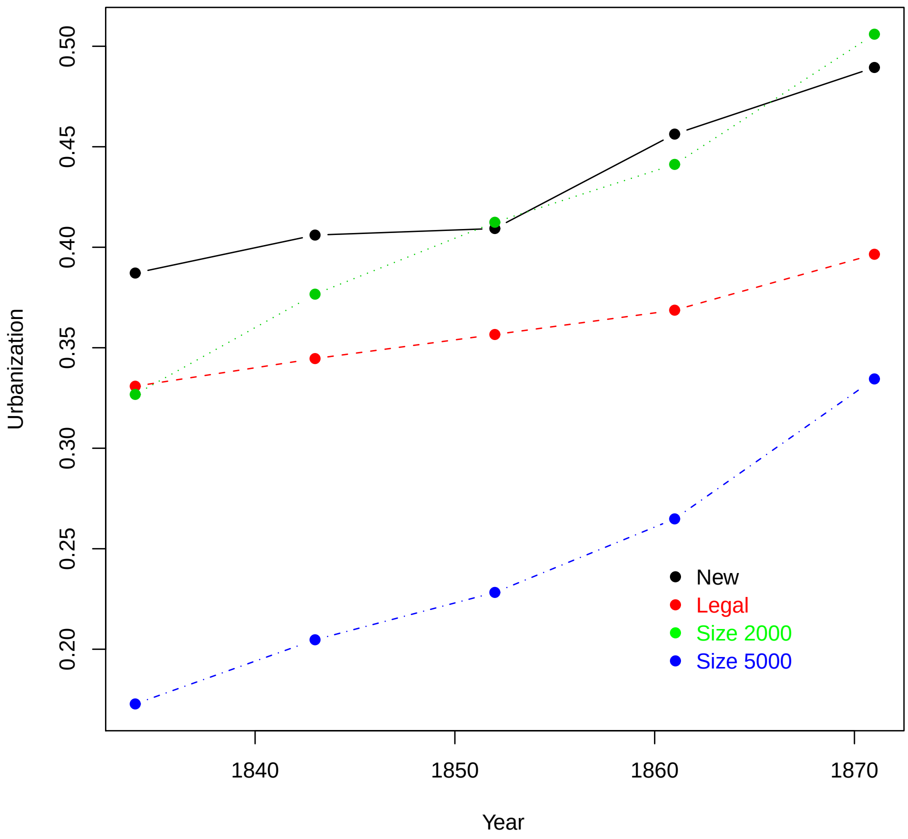
The graph depicts the p-value of the F-test determining the validity of the difference in endowment influence above and below a specific threshold value. The threshold is marked through the horizontal line.

Figure 2: Thresholds for Saxony 1834-1871



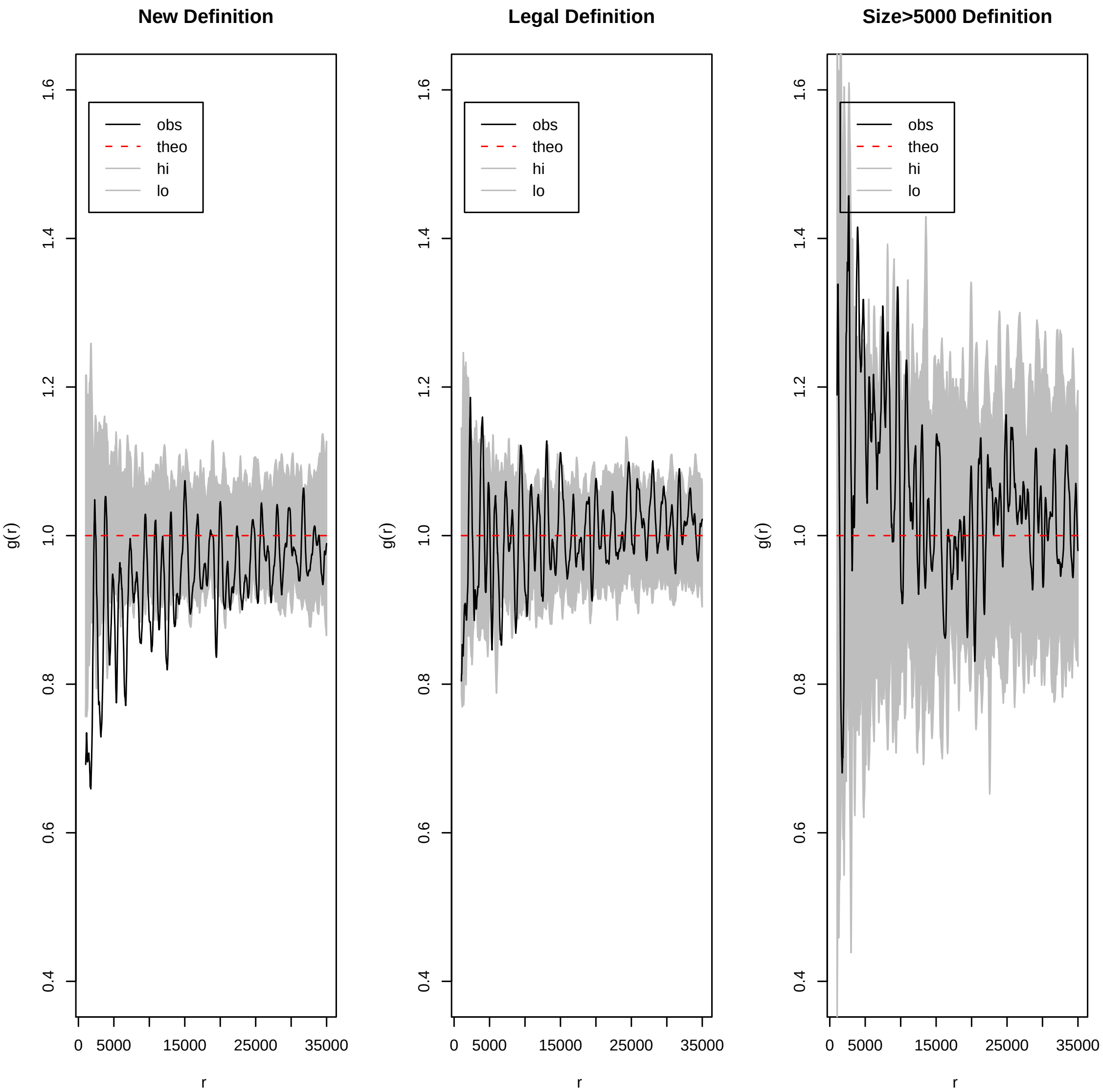
The graph is identical to Figure 1, but demonstrates the thresholds for multiple years. The respective thresholds are marked through the horizontal lines.

Figure 3: Urbanization development for different measures



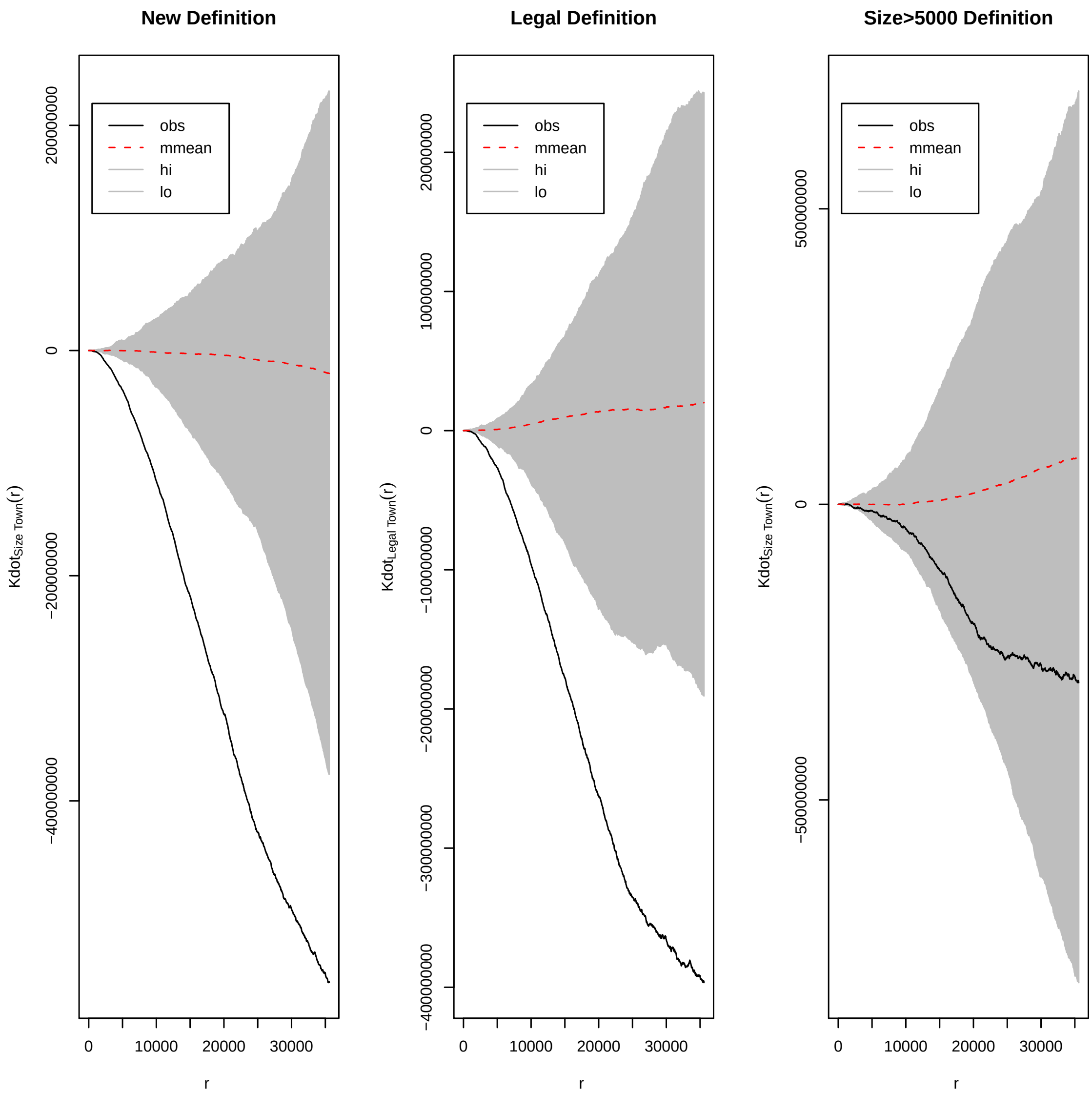
The graph depicts the temporal development of the urbanization rates implied by the different town definitions.

Figure 4: Pattern of Villages around Towns



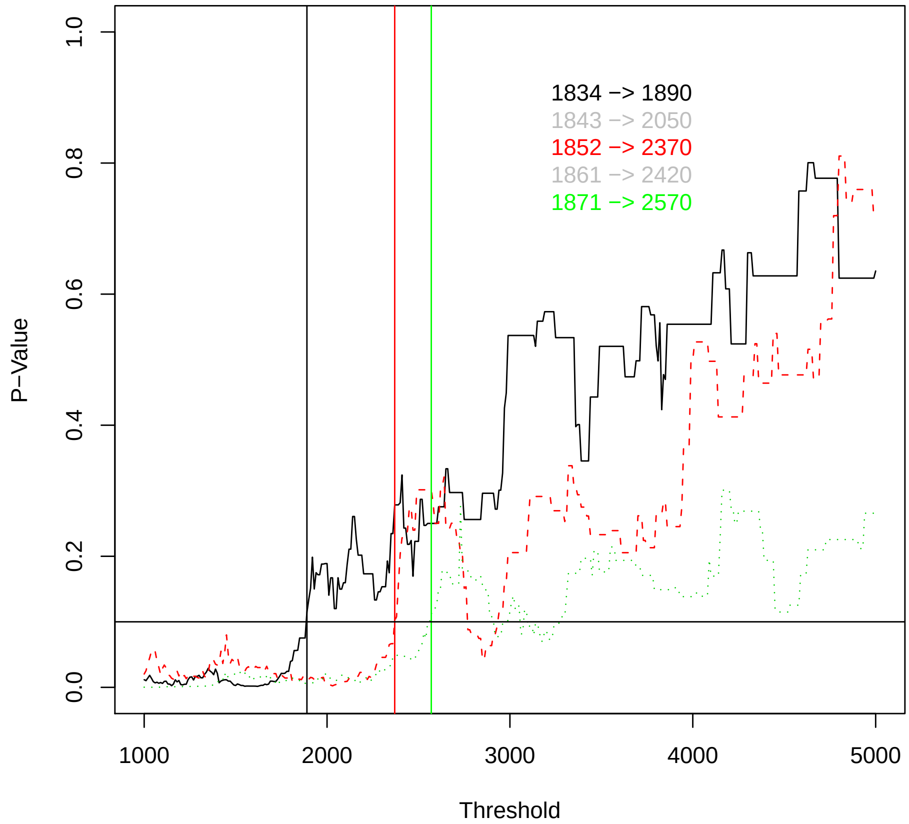
The graph depicts the function values of $g(r)$ for distance r. The value1 indicates randomness, a positive deviation implies clustering and a negative one indicates dispersion/regularity.

Figure 5: Random Labelling



The graph depicts the test for the null hypothesis that town status is spatially randomly distributed over all locations.

Figure 6: Threshold determination, alternative method



The graph depicts the p-values of the F-tests determining whether endowments have an influence on the sample above the particular threshold value for multiple years. The respective thresholds are marked through the horizontal lines.

Tables

Table 1: Influence of agricultural endowments on location size

	All Locations	Villages	Towns
(Intercept)	5.29*	4.903*	6.96*
	(0.229)	(0.206)	(0.536)
Farm	0.019*	0.02*	0.052*
	(0.005)	(0.005)	(0.019)
Pasture	0.014*	0.02*	-0.002
	(0.006)	(0.005)	(0.014)
Farm:Pasture	-0.001*	-0.001*	-0.001
	(0.000)	(0.000)	(0.000)
N	3494	3340	154
R^2	0.088	0.09	0.088

*** $p \leq 0.01$,** $p \leq 0.05$,* $p \leq 0.1$

The table reports results of standard OLS regressions of logarithmic settlement size on agriculture endowment characteristics. The samples are all locations within the study area as well as the Villages and Towns subsamples following the classification results. The observations are for the year 1834.

Table 2: Town Growth on Town Size

	New	Legal	Size>5000
Intercept	0.789***	1.098***	1.461***
	(0.121)	(0.065)	(0.247)
Population	0.037**	0.03***	0.018
	(0.015)	(0.008)	(0.012)
N	154	140	22
R^2	0.037	0.096	0.097

*** $p \leq 0.01$,** $p \leq 0.05$,* $p \leq 0.1$

The specification regresses the annualized growth rate on town size in 1834 (rescaled by 1000).

Table 3: Influence of Geography on settlement likelihood

	New		Legal		Size>5000	
Intercept	-22.9815	**	-20.1823	**	-30.1187	**
	(1.9491)		(1.8904)		(6.6230)	
Farmland Quality	0.0552		0.0296		0.1729	
	(0.0338)		(0.0314)		(0.1201)	
Pastureland Quality	0.0504		-0.0359		0.1010	
	(0.0322)		(0.0341)		(0.1302)	
Farm X Pasture	-0.0009		0.0000		-0.0029	
	(0.0006)		(0.0006)		(0.0024)	
Elevation	-0.0063	**	-0.0041	**	-0.0141	**
	(0.0014)		(0.0014)		(0.00560	
Ruggedness	0.0050		0.0068		0.0091	
	(0.0071)		(0.0079)		(0.0210)	
Temperature	-0.0488	**	-0.0124		-0.0487	
	(0.0135)		(0.0139)		(0.0529)	
Rain	0.0086	**	0.0045	**	0.0135	**
	(0.0016)		(0.0017)		(0.0053)	
Rivers	0.4860	**	0.9802	**	0.4158	
	(0.1741)		(0.1874)		(0.4759)	
River Elbe	0.0121	**	0.0074	**	0.0169	
	(0.0036)		(0.0035)		(0.0106)	
Major Roads	-0.0185	**	-0.0126		-0.0109	
	(0.0080)		(0.0080)		(0.0203)	
Minor Roads	0.0106		-0.0020		0.0007	
	(0.0087)		(0.0093)		(0.0225)	

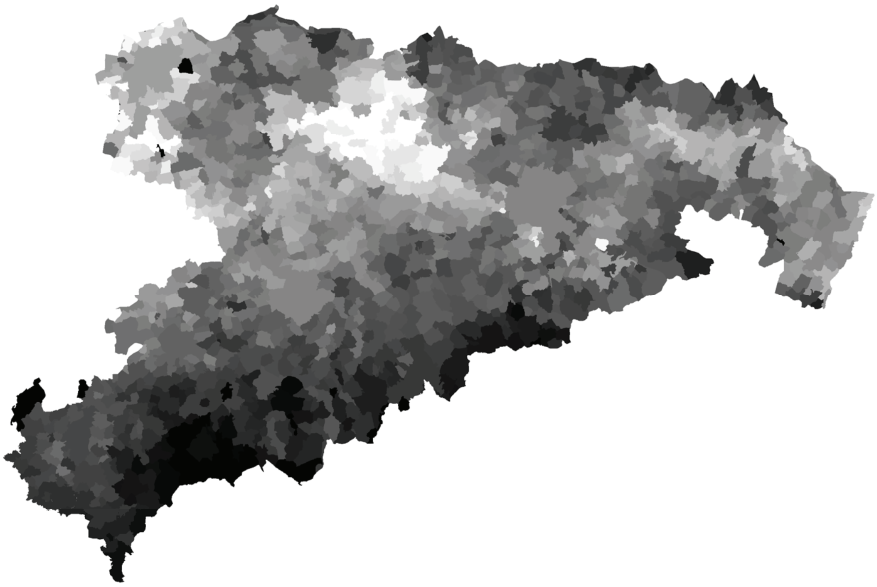
*** $p \leq 0.01$,** $p \leq 0.05$,* $p \leq 0.1$

Maps

Map 1: Farming quality values

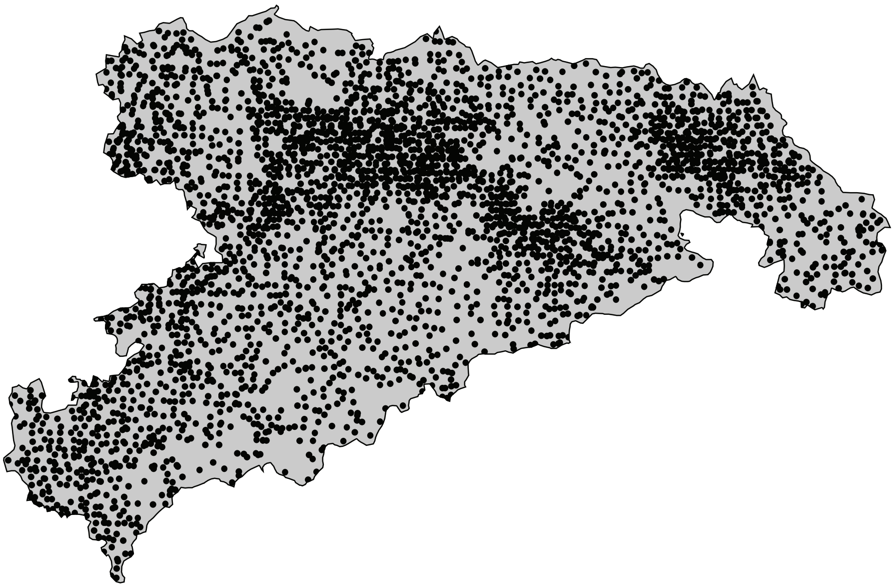
The map depicts the aptitude of land for farming purposes. Black is low quality, white high quality.

Map 2: Pasture quality values



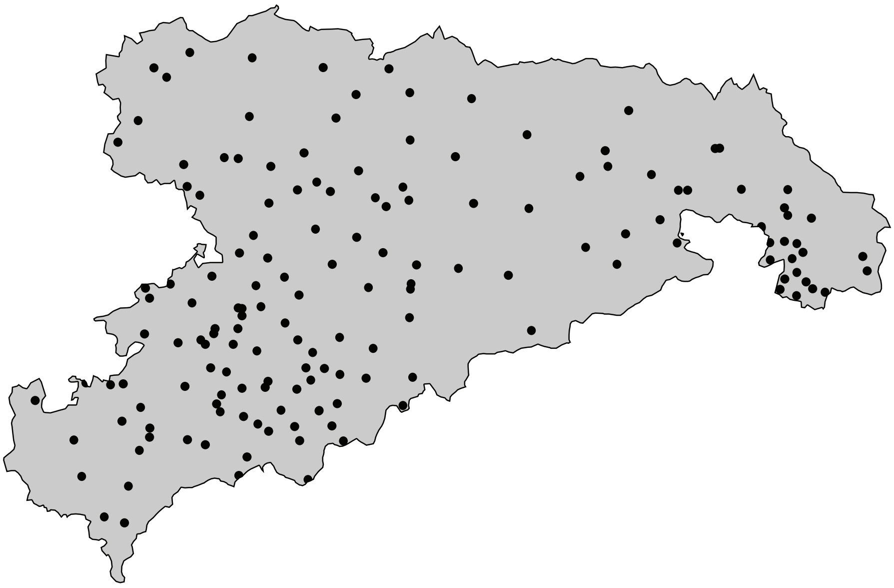
The map depicts the aptitude of land for pasture purposes. Black is low quality, white high quality.

Map 3: Settlements in Saxony 1834



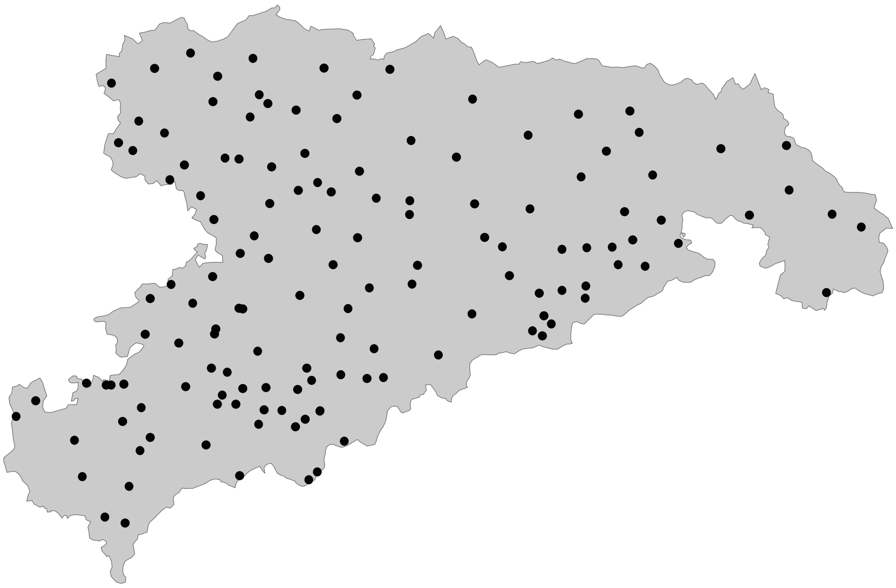
The map depicts all settlements within Saxony in its borders of 1834.

Map 4: Towns New Classification



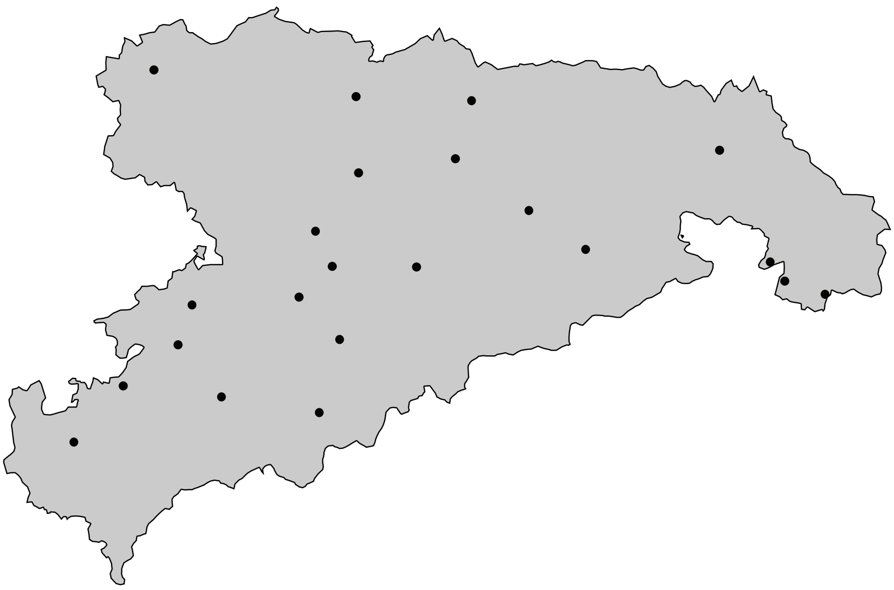
The map depicts all towns based on the new classification scheme.

Map 5: Legal Towns



The map depicts all towns based on legal status.

Map 6: Towns larger than 5000 inhabitants



The map depicts all towns based on having more than 5000 inhabitants.

Appendix

Criteria

The relevance of the two criteria, which are not used for the definition, for the historical case of Saxony is clarified. Occupational structure is one of the economic characteristics governments began to be interested in during the 19th century. Although the data isn't available for all locations, Saxony conducted an occupational census in 1849. The published results include information about the number of workers, as well as their dependents, in a number of sectors for all legal towns. The data show that 3.3% live from agriculture, 68.8% from crafts and industry,7.6% from trade and transport,9.1% work in science,art and the military,4.8% are in personal service and 6.2% are reported without regular employment (Bureau, 1854). Although these shares fluctuate from town to town, there is enough indication that these towns were characterized by a certain level of occupational diversity. The population density in settlement locations is fairly high. This is due to the historical settlement patterns, which saw clustered arrangements of houses. Furthermore some settlements erected defense arrangements, most importantly town walls, which also led to a high density of population within these walls. This implies that these two criteria can be assumed to hold for the relevant settlements. These particular characteristics were however not exclusive to Saxony, the other German states, as well as Europe, in particular western and northern parts, follow the same pattern. In general this should hold for areas where settlement processes had been concluded before the industrial revolution. The diversity of non-agricultural occupations and the density of settlements were therefore fulfilled by most if not all settlements that potentially could be identified as towns.

Alternative specification

A slightly modified approach determines the size threshold such that all observations above the threshold show no statistically significant impact of the agricultural endowments. This builds on the result that a regression of location size on agricultural endowments shows a statistical significant influence, when the full set of locations is used. In this approach the size threshold restricts the set of observations used in the estimation to locations with a population size above the threshold, while dropping those below. The resulting set of observations is then used to estimate the following specification.

$$\ln(Pop) = \alpha + \beta_f Farm + \beta_p Pasture + \beta_{fp} Farm * Pasture + \epsilon$$

The determination of the actual threshold is again based on a statistical test such that all three coefficients on agricultural endowments are jointly insignificant.

Figure 6 plots the results for this statistical test for each tested size threshold as well as the selected threshold. This results for the year 1834 in a threshold value between towns and villages of 1890. This number is fairly close to the one derived above. This results in 117 towns and a degree of urbanization of 34.5% .

Geographic Factors

The data used in this analysis is taken from (Ploeckl, 2010*b*), which is also the source for the following description of the geographic factors.

Farmland quality This variable indicates the quality of the soil with respect to farming purposes, based on public geological surveys during the middle of the 20th century. The surveys are based on thousands of measurements, and report average values for about 1600 parishes covering all of Saxony. The classification scheme uses a scale of 0-100, which is the same specification used for the empirical analysis.Saechsisches
Source: Ministerium fuer Umwelt und Landwirtschaft:GEMDAT-LABO Database, Akademie der Landwirtschaft der DDR, Muencheberg-Eberswalde

Pasture quality This variable indicates the quality of the soil with respect to pasture purposes. The data are based on the same surveys as the farmland quality and the variable is specified in the same way.
Source: Ministerium fuer Umwelt und Landwirtschaft:GEMDAT-LABO Database, Akademie der Landwirtschaft der DDR, Muencheberg-Eberswalde

Elevation This variable indicates the elevation over sea level measured in meters; the data are from current digital elevation models.
Source:U.S.Geological Survey ,National Elevation Data

Ruggedness This variable indicates the flatness of the area immediately surrounding the town. The elevation profile of an area influences agricultural suitability as well as ease of transportation. I specify this as the standard deviation of all elevation values within a two kilometer radius of the town's location.
Source:U.S.Geological Survey ,National Elevation Data

Temperature This variable indicates the suitability of a location's annual temperature with respect to agricultural purposes. The data are based on the same surveys as the farmland quality and the variable is specified in the same way.
Source: Ministerium fuer Umwelt und Landwirtschaft:GEMDAT-LABO Database, Akademie der Landwirtschaft der DDR, Muencheberg-Eberswalde

Rain This variable indicates the average rainfall at the location. The data are based on the same surveys as the farmland quality.
Source: Ministerium fuer Umwelt und Landwirtschaft:GEMDAT-LABO Database, Akademie der Landwirtschaft der DDR, Muencheberg-Eberswalde

Rivers This variable indicates whether there is a flowing water body within a kilometer of the town location, which is specified as a simple dummy variable.
Source: Saechsisches Ministerium fuer Umwelt und Landwirtschaft: -Gewässerdurchgängig (Oberflächengewässer)

Elbe One of the major means of transportation in the early 19th century was shipping, especially so on rivers. In Saxony, only the Elbe offered this possibility, as

no other river was navigable. Rivers also have other effects such as as a source of energy, but this variable captures the effect of shipping, since most Saxon towns were located at rivers. The variable is distance to the Elbe measured in km.
Source: Saechsisches Ministerium fuer Umwelt und Landwirtschaft:
-Gewässerdurchgängigkeitsprogramm (Oberflächengewässer)

Roads The data for roads are based on information from a number of historical maps. Maps drawn in 1834 show the network of major trade routes spanning Saxony and its neighbors; road classifications are quite consistent between them. Major roads either saw service by *Eilwagen*, regular priority people transportation, or were chauseed. Small roads are all other marked important road connections. The exact routes within Saxony are based on a detailed 1852 Saxony road map.